

## Association for Information Systems AIS Electronic Library (AISeL)

---

ECIS 2011 Proceedings

European Conference on Information Systems  
(ECIS)

---

Summer 10-6-2011

# TRACKING THE DIGITAL FOOTPRINTS OF CUSTOMERS: HOW FIRMS CAN IMPROVE THEIR SENSING ABILITIES TO ACHIEVE BUSINESS AGILITY

Christoph Seebach

Immanuel Pahlke

Roman Beck

Follow this and additional works at: <http://aisel.aisnet.org/ecis2011>

---

### Recommended Citation

Seebach, Christoph; Pahlke, Immanuel; and Beck, Roman, "TRACKING THE DIGITAL FOOTPRINTS OF CUSTOMERS: HOW FIRMS CAN IMPROVE THEIR SENSING ABILITIES TO ACHIEVE BUSINESS AGILITY" (2011). *ECIS 2011 Proceedings*. 258.  
<http://aisel.aisnet.org/ecis2011/258>

This material is brought to you by the European Conference on Information Systems (ECIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2011 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# **TRACKING THE DIGITAL FOOTPRINTS OF CUSTOMERS: HOW FIRMS CAN IMPROVE THEIR SENSING ABILITIES TO ACHIEVE BUSINESS AGILITY**

Seebach, Christoph, Goethe University Frankfurt, Chair of Business Administration esp.  
Information Systems, Grüneburgplatz 1, 60323 Frankfurt (Main), Germany,  
cseebach@wiwi.uni-frankfurt.de

Pahlke, Immanuel, Goethe University Frankfurt, Chair of Business Administration esp.  
Information Systems, Grüneburgplatz 1, 60323 Frankfurt (Main), Germany,  
pahlke@wiwi.uni-frankfurt.de

Beck, Roman, Goethe University Frankfurt, Chair of Business Administration esp.  
Information Systems, Grüneburgplatz 1, 60323 Frankfurt, Germany, rbeck@wiwi.uni-  
frankfurt.de

## **Abstract**

*In turbulent business environments firms must be able to respond to changes rapidly in order to operate in an economical and efficient manner. To detect such changes, sensing capabilities have to be developed to support decision makers with accurate and timely information. As is known from research in marketing, consumers' information search behavior is a valuable source to gain insights about the market's buying side. Therefore, we suggest firms include data on consumers' online information search into their IT-systems to enhance their sensing capabilities. In an empirical example we predict monthly sales numbers for two car manufacturers and demonstrate how information search data can be leveraged to improve forecasting accuracy. The results clearly show that information search based models are able to outperform existing benchmark models. Beyond sales forecasting, consumers' online information search may be analyzed for a variety of purposes to enhance the sensing capabilities of enterprises.*

*Keywords: Online Consumer Search Behavior, Business Agility, Market Sensing, Sales Forecasting.*

# 1 Introduction

Unpredictable up- and downturns of demand, rapid product and service innovation, as well as hyper-competition create a harsh and turbulent business environment for firms. Under such conditions, enterprises must be able to continuously adapt their capabilities in order to operate economically and to achieve competitive advantages (Overby et al. 2006). Enterprises which detect and adapt to environmental changes more quickly than others are characterized by higher business agility (Goldman et al. 1995). Agile firms actively sense their business environment in order to respond to such changes rapidly and to achieve superior performance (Sambamurthy et al. 2003; Day 1994). Since the sources which might be sensed and the mechanisms for response become increasingly digitized due to technological advancements, the question of how IT might help in shaping a firm's agility has gained more attention (Sambamurthy et al. 2003). Although there have been several calls to analyze the role of IT within the business agility framework (Dove 2001; Overby et al. 2006), little research has been conducted in this field so far.

Focusing on new methods to enhance the business agility of firms we analyze how IT might be used to improve market sensing through analysis of consumers' online information search (COIS). We define COIS as consumers' efforts to search for product related information online via manufacturers' or third party homepages, blogs, wikis and other social media as well as services e.g. provided by recommendation agents or search engines (Ratchford & Lee 2003; Peterson & Merino 2003). From research in marketing it is already known that information search is an excellent predictor for customers' future behavior and thus firms might be able to gain valuable information about the market's buying side (Blackwell et al. 2001; Wu & Brynjolfsson 2009). While quantifying and analyzing these search efforts has been a difficult and costly endeavor in the past (Newman & Lockeman 1975), consumers are now increasingly using the Internet when searching for information and leave their digital traces in server logs which can be used for analysis (Peterson & Merino 2003).

In this article we analyze how COIS can be used to enhance business agility and might thus improve a firm's performance. Therefore, we conduct an empirical case study based on search engine data to analyze if COIS is suitable for sales forecasting. Furthermore, we investigate the role of IT in this context, hypothesizing that exploiting COIS for business agility essentially relies on the availability of analytical information systems. Thus we contribute to the literature on business agility and consumer information search in several directions. First, we bring together these two distinct streams of research and illustrate how IT adds value to business. Moreover, to the best of our knowledge, we are the first computing out-of-sample sales for a single company based on search engine data in order to validate the prediction accuracy of our forecasting approach. In doing so, we contribute to the few existing studies on prediction in IS-literature (Shmueli & Koppius 2010).

The remainder of this paper is organized as follows. In section two we provide the theoretical background for our research questions. Therefore we define business agility as well as consumer information search and COIS, respectively. Then we bring both streams together and discuss theoretically how IT might enable business agility based on analysis of COIS. In section three we empirically discover to what extent COIS can be used to predict the sales numbers of two car manufacturers. Based on the results of our example, we then discuss how business agility of both firms might have been affected and to what extent IT would have been needed in a practical setting. Finally, we conclude with a summary of our findings and present limitations of our work as well as implications for further research.

## 2 Theoretical Background

In our proposed research approach we combine two distinct streams of literature: the strategic- and IT-management literature focusing on the business agility concept and the literature on consumer information search as an important part within marketing research. Therefore, relevant ideas from

these streams of literature are briefly highlighted in the following two sections. In the third part of this section we bring both streams together and discuss how the latter can be used to achieve the first.

## **2.1 Business Agility**

Stemming from the domain of manufacturing, business agility has been defined in several competing ways. For example, Goldman et al. (1995) and Yusuf et al. (1999) describe agility as a firm's ability to perform its business in a profitable way despite being confronted with a rapidly changing and continuously fragmenting market. This might be achieved through production of high-quality, high-performance, customer-configured goods and services (Tsourveloudis & Valavanis 2002; Sharifi & Zhang 1999). Generalizing this definition beyond the manufacturing domain, business agility has been described as a solution to maintain a firm's competitive advantage in uncertain times and turbulent business environments (Overby et al. 2006). This is achieved by the ability of enterprises to sense the market for opportunities and threats timely and to respond to such environmental changes in a rapid and effective manner (Dove 2001; Ganguly et al. 2009). In this article, we consider business agility as a firm's ability to sense and to respond to environmental changes quickly (Overby et al. 2006).

Typically, change in a firm's business environment occurs due to changes of consumer preferences, competitor's actions, regulatory or legal requirements, economic shifts, and technological progress (Sharifi & Zhang 1999). To handle such changes, enterprises should develop sensing and responding capabilities on three different dimensions of agility (Sambamurthy et al. 2003): customer agility, partnering agility, and operational agility. Customer agility reflects the competence of a firm to gain market intelligence by leveraging the voice of the customer. Moreover, partnering agility refers to a firm's competence to build effective networks with business partners and to exploit the opportunities coming along with them. Ultimately, operational agility is defined as a firm's ability to redesign and build new processes quickly in order to exploit opportunities on the market. In the following, we will focus on the first mentioned dimension: customer agility. In this regard, superior market intelligence capabilities for sensing customers' needs, and flexible product development processes to respond to them quickly have to be developed (Sambamurthy et al. 2003). As a result, enterprises may achieve above-average profitability and enjoy long-term competitive advantages (Day 1994).

Besides business capabilities, the role of IT as an important business agility enabler has also gained scientific attention (Sambamurthy et al. 2003). Although IT may affect agility directly through 'managing by wire', an enterprise's sensing and responding capabilities might be supported in an indirect way through digital options. Digital options are defined as digitized work processes and IT-systems enhancing a firm's knowledge as well as its process reach and richness. Technologies such as decision support systems provide high-quality information, enabling business units to sense their business environment timely. Moreover, processes which are better integrated increase an enterprise's boundary-spanning activities and thus enhance the ability to respond to changes in an efficient and timely manner (Overby et al. 2006). IT provides digital options, which complement an enterprise's business capabilities and thus enhance business agility and performance (Sambamurthy et al. 2003).

## **2.2 Consumer Information Search**

Within marketing science, research on consumer information search has a long history (Schmidt & Spreng 1996). Many models have been developed, most of them viewing information search as a central step within an individual's purchase decision process (Darley et al. 2010). Although varying in detail, these models consistently include the following steps: problem recognition, information search, evaluation of alternatives, purchase decision, and post-purchase (Howard & Sheth 1969; Engel et al. 1968). Having identified a consumption need consumers are often uncertain about which brand or product to choose or where to buy it. This uncertainty leads to a perceived risk of an unfavourable purchase. Therefore consumers search for information, trying to reduce their perceived risk and uncertainty about a specific purchase under consideration (Bettman 1973; Taylor 1974). Then this information is used to evaluate all alternatives before the final purchase decision is made.

Accordingly, consumers' primary goal for information search is to achieve product satisfaction or to get more value for the money spent and thus to enhance the benefits of the purchase (Punj & Staelin 1983). In doing so, consumers are collecting information from internal and external sources prior to making a decision. In this regard, internal search refers to retrieving information from memory, being gathered externally or through direct experience in the past (Schmidt & Spreng 1996). External search, in contrast, is defined as the effort to obtain environmental information from various sources, being related to the specific purchase decision to be made (Beatty & Smith 1987). To explain the amount of an individual's external search effort, the literature refers to the economics of information theory by Stigler (1961). The theory states that a consumer's information search is limited to the point where the marginal cost of additional information equal the expected benefits (Stigler 1961).

While there are many external sources to obtain information, e.g. dealer visits, print media, personal contacts or radio and TV (Moorthy et al. 1997), consumers are increasingly using the Internet when searching for information. According to a recently published study, six out of ten American adults use the Internet to search for information about the products they intend to buy online. Furthermore, 21% of Americans search for product information on the Internet every day (Jansen 2010). The Internet became the universal search instrument providing access to large amounts of information at low transaction costs (Porter 2001). While consumers obtain product information from manufacturers' or third party homepages, social media sites and recommendation agents (Peterson & Merino 2003), they start their search for information quite often using a search engine (Kumar et al. 2005).

### **2.3 The Business Value of Consumers' Online Information Search**

IT-based innovations often fundamentally change the nature of business. While this is true for other innovations as well, innovations based on IT generate vast amounts of data accessible for a variety of analytical efforts (Kohli & Grover 2008). Particularly, the Internet has dramatically changed the way business is conducted and is at the same time generating vast amounts of data. Switching from physical to online information search, consumers leave their digitized traces on the web. While quantifying and analyzing these search efforts has been a difficult and costly endeavour in the past (Newman & Lockeman 1975), there is now the opportunity to discover the recorded COIS on the web.

Specifically, it has been shown that COIS reveals valuable information about a customer's intention to purchase, even if the actual transaction is executed offline (Wu & Brynjolfsson 2009; Kumar et al. 2005). Although buying intentions may change due to unexpected circumstances, they are the best predictor for future behavior of a company's customers (Blackwell et al. 2001), and so provide an opportunity for sensing the market. Purchase intentions can be used by managers for a variety of purposes, such as sales forecasting, evaluation of advertising activities, or market segmentation (Morrison 1979). In short, monitoring COIS, firms can enhance their market sensing capabilities and thus improve their customer agility which may lead to superior business performance. The Internet allows to access COIS from search engines or from the site visits of Internet users browsing web pages. Although it may be difficult to obtain data on COIS from all possible sources, there are providers such as Google which offer search data for free download. However, how this data can be used and how IT-driven agility can be drawn from COIS is still not entirely understood. Since large amounts of raw data files have to be analyzed to extract consumer intentions from these sources, analytical Information systems such as decision support systems, data warehouses, and online analytical processing are needed for applying data mining algorithms to this kind of data (Kohli & Grover 2008; Overby et al. 2006).

From a business agility point of view, information about consumers' intentions based on analysis of COIS can be provided by IT as a digital option. Knowledge reach is enhanced through integration of this new source of data in a firm's information system. The knowledge base for decision making is enriched by using data mining technologies for pattern detection and by timely provision of high-quality information regarding customer needs. This leads to an increased market sensing capability which may enhance business agility by option. Depending on whether the business units make use of this digital option, competitive actions may follow, eventually leading to superior performance.

Consistent with findings from the IT management literature, the value of COIS for enterprises evolves from the complementarity of IT and business (Sambamurthy et al. 2003; Kohli & Grover 2008).

### 3 Sales Forecasting based on Consumer's Online Information Search – Evidence from an Empirical Example

In this section we will provide an empirical example to illustrate how companies might use search engine data to exploit COIS for business agility. Therefore, we focus on the German automotive sector which struggled heavily within the aftermath of the financial crisis in 2009. To stimulate sales, the German government decided to provide a bonus to consumers who trade in their old vehicle for a new one. The result was a big upturn in sales in 2009, followed by a big downturn in 2010. In other words, the German car manufacturers have been subject to massive environmental changes in the last couple of years. From a business agility perspective, such turbulent environments require firms effectively to sense their market in order to react swiftly to occurring changes. One of the most important capabilities of a firm which is based on market sensing is the ability to accurately forecast future sales (Day 1994; Dalrymple 1975). For instance, firms with a superior forecasting capability may be faced with lower inventory write-offs and thus achieve higher profitability (Overby et al. 2006).

According to the aforementioned line of argumentation, we were interested in analyzing the extent to which COIS may help the automotive industry to calculate monthly sales forecasts more accurately and so might improve their sensing capabilities. For this exemplary illustration, we focused on the two largest German car manufacturers VW and Opel (KBA 2010). We measured the amount of searches for both firms using search engine data from Google, since Google is the leading search property with a market share of almost 80 percent in Germany (comScore 2008).

So far, there have been only a few studies published which build upon data from Google. For example, Ginsberg et al. (2009) used Google search data to predict flu outbreaks. Other researchers analyzed the relationship between housing prices and specific Google searches and found strong correlations (Webb 2009; Wu & Brynjolfsson 2009). Finally, there are some studies analyzing the search data to predict unemployment rates (Askatas & Zimmermann 2009; Choi & Varian 2009). While the majority of the aforementioned examples use Google search data to predict macroeconomic developments, to the best of our knowledge there are no studies using Google search data to predict sales numbers on firm level. Moreover, we are the first to compute out-of-sample predictions with different prediction horizons of up to three months. Probably the most distinctive aspect of our approach is that we analyzed the value of consumer generated data through the lens of IT-driven business agility.

#### 3.1 Data Collection and Sample Description

Before we were able to start our empirical analysis, several activities had to be conducted to create our datasets as depicted in Figure 1. The remainder of this section follows these steps accordingly.

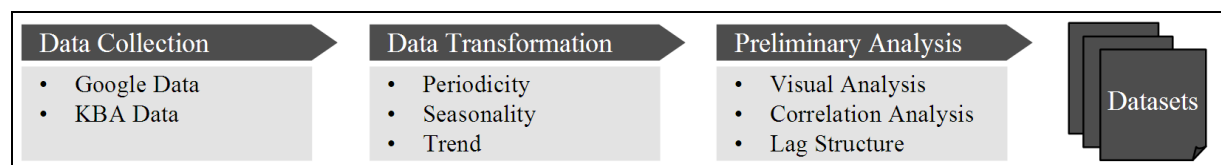


Figure 1: Data Collection and Dataset Generation Approach

First, we collected data from two sources: the monthly new vehicle registrations for Opel and VW in Germany, and the Google search volumes for both companies. Following the approach of Requena-Silvente and Walker (2007), we decided to choose the new registrations as a proxy for sales since these are available monthly, while company sales data are reported only quarterly. The registration data is issued by the German Federal Motor Transportation Authority (KBA) one week after the end of each reporting period and can be obtained separately for each car manufacturer since the beginning of

2001 (KBA 2010). As a proxy for consumers' intentions to buy a new car, we measured COIS with statistics from a service called Google Insights for Search (GIS). While other tools offer similar data on Google, e.g., Google AdSense, GIS allows access to a long history of search data on a fine-grained level for free. Reports on user queries for a specific search string are provided to the user on a weekly basis from 2004 until today (GIS 2010). Moreover, statistics on a certain query can be obtained for predefined categories such as 'vehicle shopping' and for different geographical regions. Instead of the absolute number of searches for a specific query, Google provides a search volume index, computed as the number of searches for a given term relative to the total number of searches done on Google over time. The index is scaled from 0 to 100 and may be interpreted as the likelihood of a random user to enter a specific query in a specific region and period. Thus, we had to decide on the search queries and categories to create our Google data sample. Our goal was to ensure that the search volume index is most likely to capture consumers' buying intentions. Therefore, we systematically generated weekly CSV-files for all combinations of the brand names and car models within different categories for each car manufacturer, e.g. "VW + Volkswagen + Golf + Passat + ..." in the GIS category "car purchase". Finally, we had collected 352 weekly observations for each Google file and 81 monthly registration volumes for VW and Opel from KBA between January 2004 and September 2010.

Next we had to transform the data from both sources to account for their different reporting periods as well as for market influences such as seasonality. As the registration volumes are provided on a monthly basis only, we converted the weekly Google data to match the periodicity of the KBA data. Therefore, we computed the daily search volume values for each file applying a spline smoothing method (Wahba 1990) and aggregated these values to obtain the amount for each month. First visual inspection of the data revealed strong seasonality for both kinds of data without a significant long-term trend. Since unstable and non-stationary input data may lead to spurious correlations in our empirical analysis, we computed seasonal differences on a monthly basis for all the observations  $X$  in our datasets:  $\Delta_{12}X_t = (X_t - X_{t-12})$ . That is, from each month  $X_t$  we subtracted the value of the corresponding month of the previous year  $X_{t-12}$  to obtain the deseasonalized value  $\Delta_{12}X_t$  (Chatfield 2001). For the sake of comparability we then computed growth rates dividing  $\Delta_{12}X_t$  by  $X_{t-12}$  to measure all datasets on the same scale. In the remainder of this analysis we will refer to the year-on-year growth rates for the Google search volume indices as SVI and for the new vehicle registrations as NVR.

In our preliminary analysis, we first recalled the theory on consumer information search. We expected that consumers would be likely to search for information about a car well in advance to the purchase itself. Since a vehicle is a durable good and a quite expensive investment, potential buyers should try to reduce their risk of buying an unsuitable or overpriced car through intensive information search (Armstrong et al. 2000). Plotting the SVI against the NVR, we found evidence for our assumption. Figure 2, depicts the growth rates for the search queries "Volkswagen + VW" as well as "Opel + GM + General Motors" within the Google category 'vehicle shopping' and the actual new registrations for both car manufacturers. As indicated by the dotted line, the SVI is leading the NVR and there seems to be a significant correlation between both. One can also clearly see that the governmental stimulus plan had the looked-for effect illustrated by increasing sales numbers in 2009.

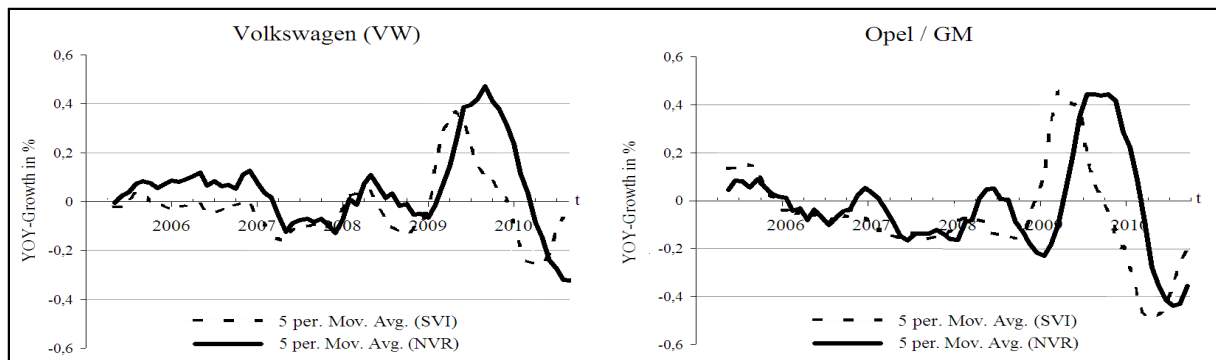


Figure 2: Plots of SVI and NVR Growth Rates for Volkswagen and Opel between 07.2005 and 09.2010

Next we conducted an analytical approach to explore the data for the best combination of search terms and categories. Therefore we computed correlations between the transformed values of each Google-file and the new registrations for time-lags between 1 to 6 months. For both car manufacturers we found strongest correlations for exactly the same combination of factors, signalling validity of our preliminary results. Simple combinations of the brand names without specific car models “Volkswagen + VW” and “Opel + GM + General Motors” within the Google category ‘vehicle shopping’ with a time-lag of 3-4 months lead to a correlation coefficient of more than 0.7. While we discovered the correlation coefficients for lag 1 to 6, we identified a figure similar to normal distribution. Unfortunately, the data did not allow us to test if the duration of consumers’ search behavior for both car manufacturers is normal distributed. Nevertheless, it seems reasonable to assume that some customers start searching earlier and some later than the average when buying a new car. Therefore, we computed a weighted moving average  $SVI_{vw/opel_{t-1,2,3,4,5,6}}$  for both car manufacturers, by weighting the values of the preceding 6 months with factors obtained through the assumption of normal distribution. Computing the correlations again the coefficients rose to a value of 0.8 (Opel 0.81; VW 0.79) for both car manufacturers. As these two datasets  $SVI_{vw/opel_{t-1,2,3,4,5,6}}$  were leading the NVR by only one month, we created two additional datasets for each car manufacturer to enhance our prediction horizon. Therefore, we set the weights for  $t-1$  and  $t-6$  to zero to create the datasets  $SVI_{vw/opel_{t-2,3,4,5}}$  which are suitable for 2-months ahead predictions. We then applied the same method to obtain the datasets  $SVI_{vw/opel_{t-3,4}}$ .

For further analysis, we decided to keep these three SVI datasets for each car manufacturer. Due to necessary data preparation measures we lost 12 observations computing the growth rates and another 6 observations for calculating moving averages. Thus, the initial datasets were reduced by 18 observations and finally consisted of 63 data points for each car manufacturer.

### 3.2 Statistical Models

For our empirical analysis we first defined three simple linear regression models for each car manufacturer to measure the relationship between the NVR as dependent and the SVI as the independent variables. We then used the same models to test the predictive power of the SVI computing in-sample and out-of sample predictions for different horizons.

$$\textbf{SVI-model 1a:} \quad NVR_{vw/opel_t} = \alpha + \beta SVI_{vw/opel_{t-1,2,3,4,5,6}} + \varepsilon$$

$$\textbf{SVI-model 1b:} \quad NVR_{vw/opel_t} = \alpha + \beta SVI_{vw/opel_{t-2,3,4}} + \varepsilon$$

$$\textbf{SVI-model 1c:} \quad NVR_{vw/opel_t} = \alpha + \beta SVI_{vw/opel_{t-3,4}} + \varepsilon$$

We defined two well-accepted benchmark models based on past registration data for the sake of an easy comparison with our models (Armstrong et al. 2000; Makridakis et al. 1979). While there are many models one could choose for this purpose, we decided to apply these models not only because they are easily comprehensible and simple to apply. Moreover, predictions based on extrapolation of past data are also commonly used by managers in practice (Mentzer & Kahn 1995).

The first is a simple autoregressive model (Chatfield 2001) estimating the NVR in  $t$  based on the lagged NVR in  $t-1$ :

$$\textbf{B-Model A:} \quad NVR_{vw/opel_t} = \alpha + \beta NVR_{vw/opel_{t-1}} + \varepsilon$$

The second model we used is the naive model, which is also referred to as the random walk. The model uses the most recent observation as a forecast. That is the growth rate of the last month is assumed to be the growth rate of the next month. Despite its simplicity the model usually produces good results and has often been suggested as a benchmark in the literature (Makridakis et al. 1982).

$$\textbf{B-Model B:} \quad NVR_{vw/opel_t} = NVR_{vw/opel_{t=\text{last Observation}}}$$

Assessing the predictive accuracy of our models to make them comparable, we examined the different prediction error measures discussed in the literature (Armstrong & Collopy 1992; Makridakis et al. 1979; Hyndman & Koehler 2006). According to Hyndman et al. (2008), we chose to compute the



mean absolute error (MAE) for all our predictions. The MAE measures the error in the same unit as the target value and is easy to interpret. Since our goal is to compare different forecasting methods within each time-series a scale-dependent measure such as the MAE is appropriate for our purpose (Hyndman et al. 2008). Additionally, we decided to compute the root mean square error (RMSE) as it is more sensitive to outliers than the MAE. Moreover, both measures are closely related to decision making, which is the ultimate purpose of sales forecasting (Armstrong & Collopy 1992).

### 3.3 Analysis and Results

To evaluate the explanatory power of our models, we first ran in-sample OLS-regressions for all SVI-models and each car manufacturer as well as for our autoregressive benchmark model. Overall, we found a strong and significant relationship for all our SVI-models. As it is displayed in Table 1,  $R^2$  values are between 0.5 and 0.7 and all coefficients are significant at  $p < 0.01$ . Compared to the values obtained for benchmark model A, all SVI-models are stronger related to our dependent variable. To control for lagged effects, we combined the autoregressive model A with each SVI-model to test if the Google data helps to enhance the explanatory power of this baseline model. As it can be seen from the results in brackets, the  $R^2$  of model A increases when augmented with one of the SVI-models. Thus we conclude that the Google data provides intrinsic explanatory power which cannot be explained by a lagged value of our dependent variable. We went further and used the results to predict monthly sales growth rates for each manufacturer from in-sample. The results in Table 1 reveal that all models based on the SVI significantly outperform the benchmark models when compared by the MAE and RMSE. In fact, the best SVI-based models for each car manufacturer predict the dependent variable by an accuracy of below 11 percent compared to the absolute actual value.

	Volkswagen (VW)					Opel / GM				
	SVI-models			Benchmark models		SVI-models			Benchmark models	
Independent variable	1a	1b	1c	A	B	1a	1b	1c	A	B
SVIvw/opel $t-1,2,3,4,5,6$	1.49** (1.53**)	-	-	-	-	1.37** (0.96**)	-	-	-	-
SVIvw/opel $t-2,3,4,5$	-	1.65** (1.65**)	-	-	-	-	1.65** (1.14**)	-	-	-
SVIvw/opel $t-3,4$	-	-	2.44** (1.96**)	-	-	-	-	2.21** (1.73**)	-	-
NVRvw/opel $t-1$	(-0.02)	(-0.03)	(0.2)	0.63**	-	(0.19*)	(0.18)	(0.23**)	0.7**	-
$R^2$	<b>0.62</b> (0.62)	0.58 (0.58)	0.53 (0.55)	0.39	-	0.66 (0.68)	<b>0.58</b> (0.68)	0.64 (0.67)	0.49	-
MAE	<b>10.73</b>	11.36	12.03	13.48	12.09	11.00	<b>10.71</b>	10.97	14.11	17.13
RMSE	<b>13.50</b>	14.22	14.97	17.17	16.04	14.60	<b>14.59</b>	14.99	17.86	22.11

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; Bold values indicate best model with lowest error; values in brackets for lag-control

Table 1: Linear Regression Results on NVR and SVI for Volkswagen and Opel

So far, our in-sample analysis has demonstrated that models based on Google data are able to outperform well-accepted benchmark models. However, from the literature we know that accurate in-sample predictions are not a guarantee for a good performance outside the sample period (Tashman 2000, Shmueli & Koppius 2010). Therefore, we computed out-of-sample predictions for the last 12 months using a rolling window. That is, we first recursively estimated the parameters for our models, leaving out the last 12 observations in the case of a 1-month ahead prediction. Then, we predicted the value of the earliest month that was omitted and re-estimated the parameters leaving out the last 11 observations for the next prediction and so forth. Testing the predictive power of our models for different time horizons, we analogously applied the same procedure for 2- and 3-month ahead predictions. Then we computed the MAE as well as the RMSE over all 12 predictions for each prediction horizon. The results displayed in Table 2 for Volkswagen and Table 3 for Opel confirm the strong predictive power of the Google data, found in our in-sample estimations. Again, the errors of our 1-month ahead predictions are significantly lower for the Google-based forecasts compared to the

benchmark. Moreover, we expanded our prediction horizon from one up to three months and the error of the SVI-based models remained to be the lowest in any case.

Volkswagen (VW)									
Pred. horizon (months)	1			2			3		
Model	SVI	Benchmark		SVI	Benchmark		SVI	Benchmark	
	1a	A	B	1b	A	B	1c	A	B
MAE	<b>11.60</b>	12.09	15.45	<b>13.38</b>	14.98	16.56	<b>15.71</b>	20.56	21.58
RMSE	<b>13.24</b>	16.04	17.42	<b>14.92</b>	17.58	18.56	<b>17.32</b>	26.11	24.36

Bold values highlight model with lowest error

Table 2: Out-of-Sample Prediction Accuracy Volkswagen

Opel / GM									
Pred. horizon (months)	1			2			3		
Model	SVI	Benchmark		SVI	Benchmark		SVI	Benchmark	
	1a	A	B	1b	A	b	1c	A	B
MAE	<b>14.55</b>	17.13	16.34	<b>14.59</b>	25.30	24.06	<b>14.66</b>	29.49	28.34
RMSE	<b>20.64</b>	22.11	22.03	<b>20.94</b>	30.78	28.80	<b>22.12</b>	26.80	33.00

Bold values highlight model with lowest error

Table 3: Out-of-Sample Prediction Accuracy Opel/GM

Most notably, the SVI-model's performance remains comparatively stable when computing predictions with an expanded prediction horizon. While the difference between the errors of the models is quite low in the case of 1-month ahead predictions, it steadily increases in favour of the SVI-models and tops at the 3-month ahead window as depicted in Figure 3.

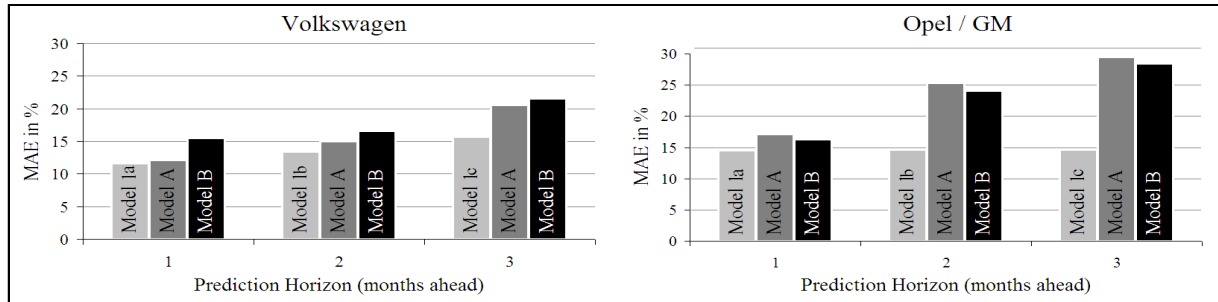


Figure 3: Prediction Error (MAE) Comparison between Models 1 (SVI) and Models A/B (Benchmark)

### 3.4 Discussion

Within the previous sections we have demonstrated how COIS can be used to compute sales forecasts. Moreover, the results of our empirical investigation clearly illustrate that sales predictions based on COIS are able to outperform well-accepted benchmark models. No matter which of the 3 prediction horizons we chose, the forecasts of the SVI-models were always more accurate than those of the autoregressive and the naïve benchmark models. More specifically, our analysis has shown how VW and Opel could use Google search data to improve their sales forecasting capabilities. This, in turn, might enable both firms better to anticipate their customers' demand and better to align their operative and strategic planning activities. As a result, lower inventory costs or higher efficiency may be achieved and specific marketing activities may be initiated at the right time to stimulate demand when faced with a downturn (Davis & Mentzer 2007). More generally, by analyzing COIS both car manufacturers may enhance their market sensing abilities, enabling them to faster respond to changes in their business environment. Accordingly, our empirical investigations show how enterprises such as VW or Opel could enhance their business agility using search engine data.

However, although the empirical results clearly verify our propositions, our example is rather simple, since our goal was to provide first evidence for our theoretical and practical statements. Firms have

access to additional sources of internal and external data and probably apply much more sophisticated methods to predict sales than our benchmark models. Moreover, computing forecasts based on online search data manually - as in our case - is too costly and too slow. According to the business agility concept, firms in turbulent environments need real-time data on demand to be able to act on changes in a timely manner (Overby et al. 2006). Therefore, data on COIS should be integrated into existing methods to complement and extend them. According to Armstrong et al. (2001) as well as Makridakis and Hibon (2000) such combinations often improve the accuracy of forecasts. Moreover, since data on COIS is digitized already it can probably be integrated into existing IT-systems in an automated fashion. Applying data mining methods, the data can be analyzed and combined with other information to produce high quality information timely. Users from business units might then be able to obtain customized information about the market on demand. In other words, leveraging their IT to digitize the manual process in our example, the two car manufacturers might be able to enhance their knowledge reach and richness. The business units can utilize this digital information by option, which might lead to a better market sensing ability and enhanced business agility (Overby et al. 2006).

## 4 Conclusion

Contributing to the literature on business agility (Sambamurthy et al. 2003; Overby et al. 2006), this research theoretically and practically analyzed how firms might exploit COIS to enhance their ability to sense their business environment in turbulent times. In our case, we predicted out-of-sample sales based on online search data and found strong support for our assumption that COIS is a valuable source to enhance forecasting accuracy and thus a firm's market sensing capability (Day 1994; Dalrymple 1975). Moreover, we argue that the use of IT is essential to the use of COIS, since the data has to be obtained, analyzed and provided on demand and preferably in real-time to meet the requirements of the business agility concept. Therefore, information systems are needed which integrate and analyze these masses of data automatically, offering the results as a digital option to decision makers (Overby et al. 2006). In doing so, business agility will be enhanced and business units can initiate competitive actions in a better way to achieve superior performance and business value, respectively (Sambamurthy et al. 2003). Thus, we additionally contribute to the discussion of how IT creates business value and provide evidence for the position that the value of IT arises indirectly through complementarity with business.

Our argumentation relies on the assumption that firms have access to accurate online consumer search data in a timely manner. We are aware of the fact that such data may not be obtainable for all commercial areas or industry sectors. Moreover, consumers might use other sources than those available for firms to gather information on the Internet. In this regard, a consumer's personal characteristics might have an impact on the choice of the search channel. This may lead to some kind of empirical bias, since the data available is not a representation of all consumers' search efforts. However, firms can access the data from Google which allowed us to make accurate sales predictions. Moreover, these data may be enriched by search logs from a firm's own website or any other relevant data available in a firm's IT-system. Furthermore, we are aware of the fact that our data sample is quite small for the kind of analysis we have performed. Unfortunately, the Google data is only available from 2004 on and we have lost some observations during data preparation. Since sales data are typically available on a quarterly basis we have tried to circumvent this issue by selecting the new car registrations as our dependent variable. Although we believe this is a fairly good proxy for the sales numbers, this, as well as our data preparation measures, will add some bias to our results.

Bringing together two formerly distinct streams of the literature we may have opened the door for a wide range of topics and questions for further research. From the management as well as IT management viewpoint it should be analyzed what other sensing activities beyond sales forecasting might be enhanced by analysis of COIS. Moreover, it should be investigated which other sources of online consumer search data are valuable with regard to the business agility concept and which source or combination of sources is the best choice for which sensing activity and vice versa. From a more technical perspective, integration of the data in an existing information system to achieve highly-

available and on demand accessibility might be of interest just as the inclusion in existing procedures for knowledge generation. Furthermore, other aspects to be analyzed stem from the marketing literature on consumer search behavior (Darley et al. 2010), e.g., a consumer's individual or social characteristics like age, gender or culture might have an impact on the amount of search and thus are worth considering when designing methods to exploit COIS for sensing the environment.

## References

- Armstrong, J.S., Adya, M. and Collopy, F. (2001). Rule-based forecasting: Using judgment in time-series extrapolation. *Int. Series in Operations Research and Management Science*, 259–284.
- Armstrong, J.S. and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8, 69–80.
- Armstrong, J.S., Morwitz, V.G. and Kumar, V. (2000). Sales forecasts for existing consumer products and services: Do purchase intentions contribute to accuracy? *International Journal of Forecasting*, 16(3), 383–397.
- Askatas, N. and Zimmermann, K.F. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, 55(2), 107–120.
- Beatty, S. and Smith, S. (1987). External Search Effort: An Investigation Across Several Product Categories. *Journal of Consumer Research*, 14(1), 83–95.
- Bettman, J. (1973). Perceived Risk and Its Components: A Model and Empirical Test. *Journal of Marketing Research*, 10(2), 184–190.
- Blackwell, R., Miniard, P. and Engel, J. (2001). *Consumer Behavior*. 9th Edition. Mason, OH.
- Chatfield, C. (2001). *Time-Series Forecasting*. Chapman and Hall, Boca Raton FL.
- Choi, H. and Varian, H. (2009). Predicting Initial Claims for Unemployment Insurance Using Google Trends. Technical Report, Google.
- comScore (2008), comScore Releases July 2008 German Search Rankings. Available at: [http://www.comscore.com/Press\\_Events/Press\\_Releases/2008/09/Top\\_German\\_Search\\_Engines](http://www.comscore.com/Press_Events/Press_Releases/2008/09/Top_German_Search_Engines) [Accessed March 12, 2011].
- Dalrymple, D. (1975). Sales forecasting: Methods and accuracy. *Business Horizons*, 18, 69–73.
- Darley, W.K., Blankson, C. and Luethge, D.J. (2010). Toward an integrated framework for online consumer behavior and decision making process: A review. *Psychology and Marketing*, 27(2), 94–116.
- Davis, D.F. and Mentzer, J.T. (2007). Organizational factors in sales forecasting management. *International Journal of Forecasting*, 23 (3), 475–495.
- Day, G. (1994). The Capabilities of Market-Driven Organizations. *Journal of Marketing*, 58(4), 37–52.
- Dove, R. (2001). *Response ability: the language, structure, and culture of the agile enterprise*. Jon Wiley and Sons, New York.
- Engel, J.F., Kollat, D.T. and Blackwell, R.D. (1968). *Consumer Behavior*. Holt, Rinehart & Winston, New York.
- Ganguly, A., Nilchiani, R. and Farr, J.V. (2009). Evaluating agility in corporate enterprises. *International Journal of Production Economics*, 118(2), 410–423.
- GIS (2010). Google Insights for Search, Available at: <http://www.google.com/insights/search/?hl=en-US#> [Accessed November 18, 2010].
- Goldman, S., Nagel, R. and Preiss, K. (1995). *Agile Competitors and Virtual Organizations: Strategies for Enriching the Customer*. Van Nostrand Reinhold, New York.
- Howard, J.A. and Sheth, J.N. (1969). *The Theory of buying behavior*. Jon Wiley and Sons, New York.
- Hyndman, R.J. and Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Hyndman, R.J., Koehler A.B., Ord, K.D. and Snyder, R.D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer, Berlin Heidelberg.
- Jansen, J. (2010). Online Product Research, PEW Research Center. Available at: <http://pewinternet.org/Reports/2010/Online-Product-Research.asp> [Accessed November 14, 2010].

- KBA (2010). Federal Motor Transport Authority, Available at: <http://www.kba.de/EN> [Accessed November 18, 2010].
- Kohli, R. and Grover, V. (2008). Business value of IT: an essay on expanding research directions to keep up with the times. *Journal of the Association for Information Systems*, 9(1), 23–39.
- Kumar, N., Lang, K.R. and Peng, Q. (2005). Consumer Search Behavior in Online Shopping Environments. *e-Service Journal*, 3(3), 87–102.
- Makridakis, S. et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1(2), 111–153.
- Makridakis, S. and Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4), 451–476.
- Makridakis, S., Hibon, M. and Moser, C. (1979). Accuracy of Forecasting: An empirical Investigation. *Journal of the Royal Statistical Society*, 142(2), 97–145.
- Mentzer, J.T. and Kahn, K.B. (1995). Forecasting technique familiarity, satisfaction, usage, and application. *Journal of Forecasting*, 14(5), 465–476.
- Moorthy, S., Ratchford, B.T. and Talukdar, D. (1997). Consumer information search revisited: Theory and empirical analysis. *Journal of Consumer Research*, 23(4), 263–277.
- Morrison, D. (1979). Purchase Intentions and Purchase Behavior. *Journal of Marketing*, 43(2), 65–74.
- Newman, J.W. and Lockeman, B.D. (1975). Measuring prepurchase information seeking. *Journal of Consumer Research*, 2(3), 216–222.
- Overby, E., Bharadwaj, A. and Sambamurthy, V. (2006). Enterprise agility and the enabling role of information technology. *European Journal of Information Systems*, 15(2), 120–131.
- Peterson, R.A. and Merino, M.C. (2003). Consumer information search behavior and the internet. *Psychology and Marketing*, 20(2), 99–121.
- Porter, M. (2001). Strategy and the internet. *Harvard Business Review*, 63–78.
- Punj, G. and Staelin, R. (1983). A Model of Consumer Information Search Behavior for New Automobiles. *Journal of Consumer Research*, 9(4), 366–380.
- Ratchford, B. and Lee, M. (2003). The Impact of the Internet on Information Search for Automobiles. *Journal of Marketing Research*, 40(2), 193–209.
- Requena-Silvente, F. and Walker, J. (2007). The impact of exchange rate fluctuations on profit margins: The UK car market, 1971-2002. *Journal of Applied Economics*, 10(1), 213–235.
- Sambamurthy, V., Bharadwaj, A. and Grover, V. (2003). Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms. *MIS Quarterly*, 27(2), 237–263.
- Schmidt, J.B. and Spreng, R.A. (1996). A proposed model of external consumer information search. *Journal of the Academy of Marketing Science*, 24(3), 246–256.
- Sharifi, H. and Zhang, Z. (1999). A methodology for achieving agility in manufacturing organisations: An introduction. *International Journal of Production Economics*, 62(1-2), 7–22.
- Shmueli, G. and Koppius, O. (2010). Predictive Analytics in Information Systems Research. Robert H. Smith School Research Paper No. RHS 06-138. Forthcoming in *MIS Quarterly*.
- Stigler, G. (1961). The Economics of Information. *The Journal of Political Economy*, 69(3), 213–225.
- Tashman, L.J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), 437–450.
- Taylor, J.W. (1974). The role of risk in consumer behavior. *The Journal of Marketing*, 38(2), 54–60.
- Tsourveloudis, N.C. and Valavanis, K.P. (2002). On the measurement of enterprise agility. *Journal of Intelligent and Robotic Systems*, 33(3), 329–342.
- Wahba, G. (1990). Spline models for observational data. *Society for Industrial and Applied Mathematics, SIAM*.
- Webb, G.K. (2009). Forecasting US Home Foreclosures with an Index of Internet Keyword Searches. *AMCIS 2009 Proceedings*, San Francisco.
- Wu, L. & Brynjolfsson, E. (2009). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities. *ICIS 2009 Proceedings*, Phoenix.
- Yusuf, Y., Sarhadi, M. and Gunasekaran, A. (1999). Agile Manufacturing: The Drivers, concepts and attributes. *International Journal of Production Economics*, 62(1–2), 33–43.